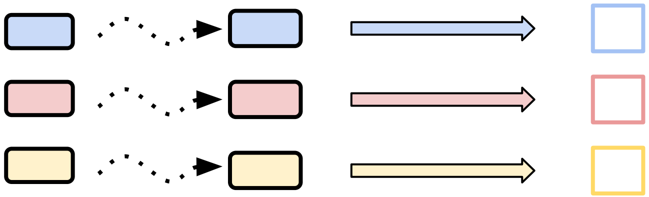
A Smoothed Analysis of the Greedy Algorithm for Linear Contextual Bandits

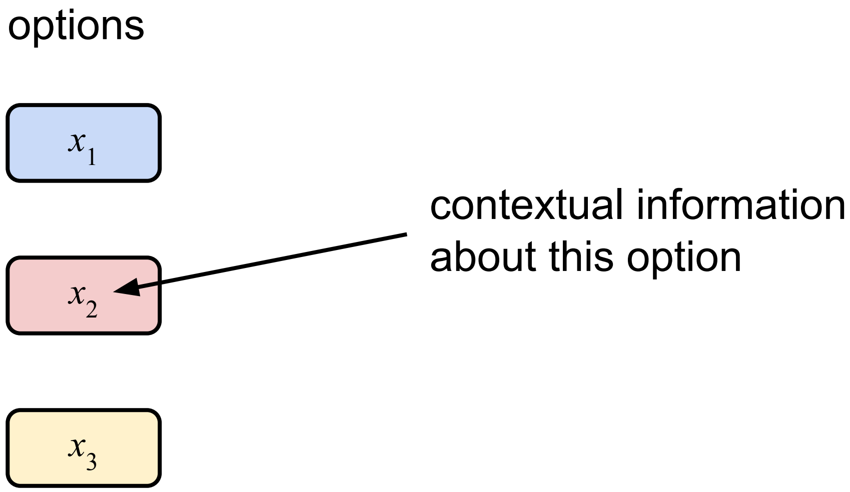


Sampath Kannan — University of Pennsylvania
Jamie Morgenstern — Georgia Tech
Aaron Roth — University of Pennsylvania
Bo Waggoner — **Microsoft Research, NYC**
Z. Steven Wu — University of Minnesota

Neural Information Processing Systems, December 2018

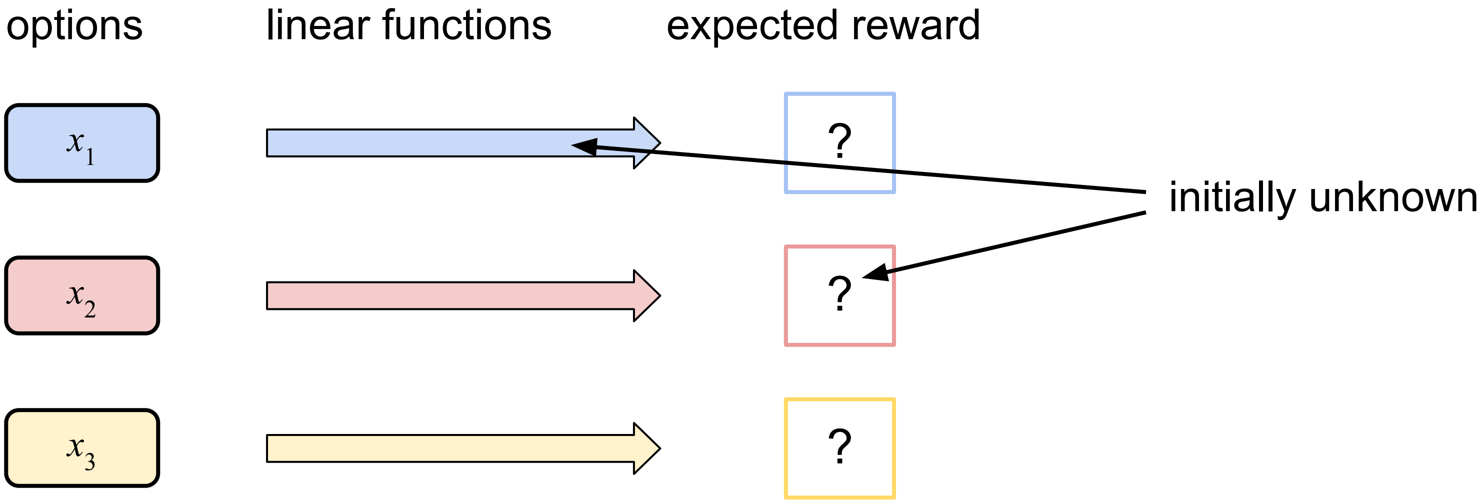
Linear contextual bandits

Model for repeated decisionmaking:



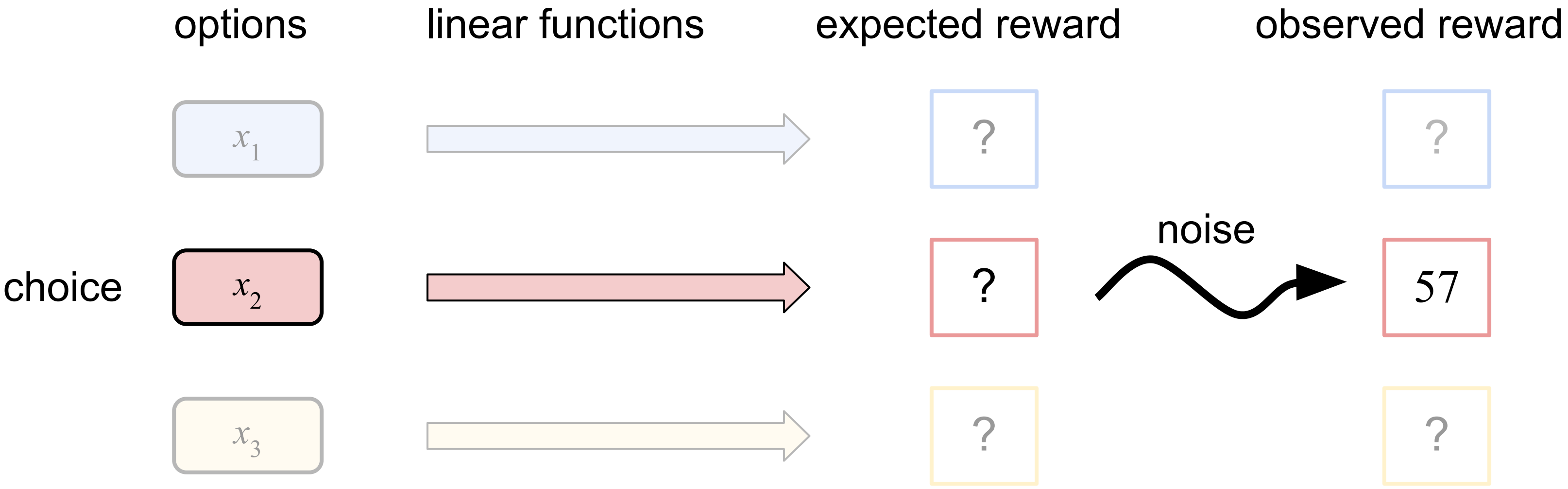
Linear contextual bandits

Model for repeated decisionmaking:



Linear contextual bandits

Model for repeated decisionmaking:

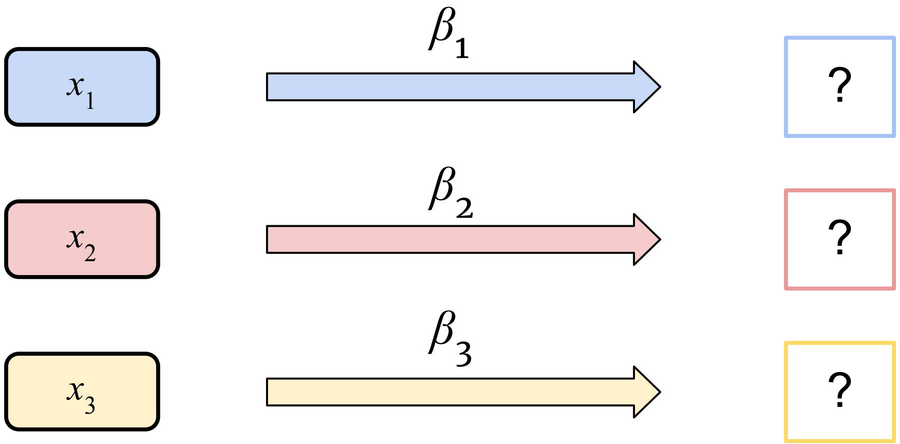


Greedy algorithm

Each step: max estimated reward **(pure exploitation)**

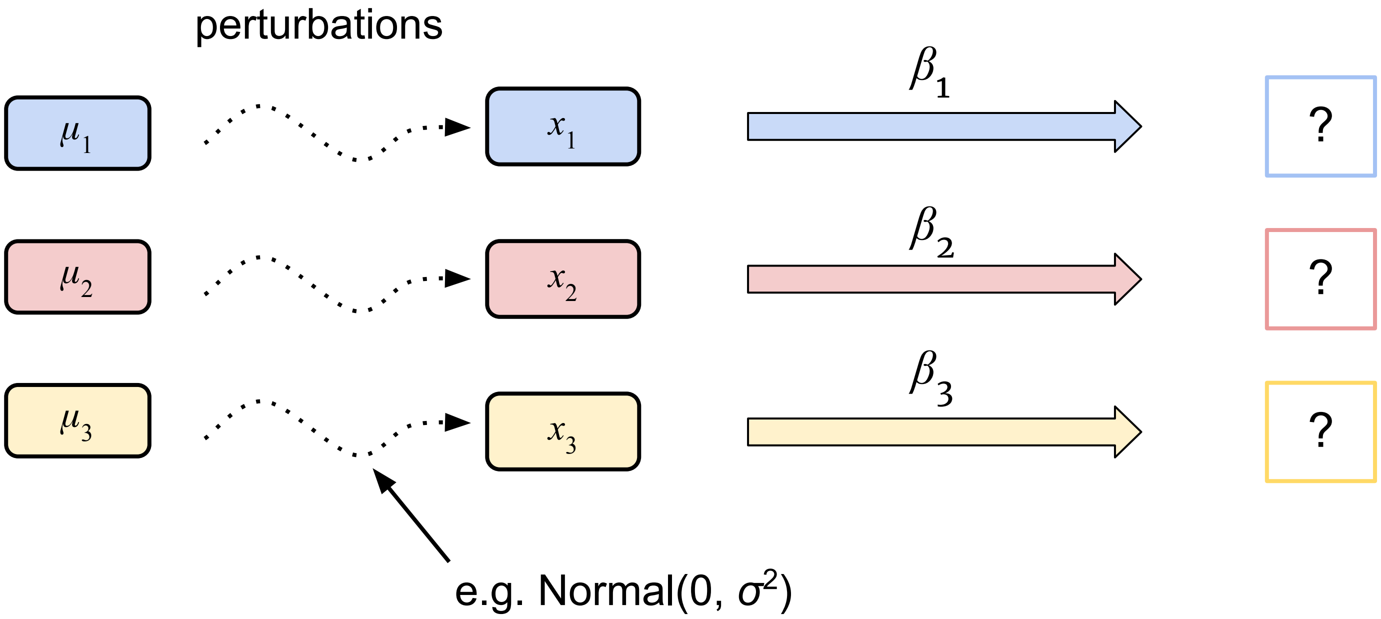
In the **worst case**: arbitrarily bad performance!

$\Rightarrow$ *Exploration seems necessary...*



Smoothed Analysis

Suppose there is some randomness in the world...



Results

Theorem. With a small amount of training data,

the Greedy algorithm achieves good performance.

Builds on Bastani, Bayati, Khosravi (2017).

Results

Theorem. With a small amount of training data,

$$n = \mathrm{poly}(1/\sigma\ ,\ 1/\min_i \| \beta_i \|\ ),$$

the Greedy algorithm achieves good performance.

$$\mathrm{Regret} \leq O(\sqrt{T})$$

Results

Theorem. With a small amount of training data,

$$n = \text{poly}(1/\sigma\ ,\ 1/\min_i \|\beta_i\|),$$

the Greedy algorithm achieves good performance.

$$\text{Regret} \leq O(\sqrt{T})$$

Theorem. In the single parameter setting ($\beta_i = \beta$), with *no initial training data*, Greedy achieves

$$\text{Regret} \leq O(\sqrt{T})$$

Motivation and future work

(1) Understand *when exploration is necessary*

(2) Understand *myopic decisionmaking*:

- Incentives
- Fairness/ethics (medical treatments)

Thanks!